

« L'hétérogénéité des corpus : approches sémiotiques ; dix-huit définitions du gène » dans *Les Cahiers de jérigo*, n°4, *Du traitement du discours dans les recherches en communication*, Bernard Castagna, dir., Université de Tours, 17p., 2001.

DIX-HUIT DÉFINITIONS SUR LE « GÈNE »

JACQUES FONTANILLE

Université de Limoges

Institut Universitaire de France

PROBLÉMATIQUE

La revue *La Recherche* a publié récemment une série de dix-huit définitions du gène, dans un article ¹ qui a eu un grand retentissement, car les lecteurs (cultivés) de cette revue ont pu se rendre compte que les plus grands spécialistes qui ont en charge le fondement scientifique des technologies actuelles et futures du vivant ne s'accordent pas du tout sur la définition de leur objet.

Malgré sa spécificité et sa technicité, le corpus est ici parfaitement typique de l'hétérogénéité des corpus d'études en communication, et notamment des corpus de presse scientifique :

- l'information est relayée par un sujet d'énonciation délégué par l'organe de presse, qui intervient plus ou moins discrètement (ici, discrètement, mais visiblement, par un chapeau introductif indiquant dans quelles conditions les définitions ont été recueillies, et par des encadrés techniques, explicatifs et à caractère didactique, pour aider le lecteur à comprendre le texte des définitions ;

- elle est déjà mise en forme par d'autres sujets d'énonciations, les "informateurs" interrogés, et qui sont en grand nombre, et d'opinion très divergente ;

- elle est partiellement "multimodale" (et syncrétique !) puisque l'article associe des textes, des schémas, et des dessins d'humour.

Dans les limites de cette analyse, mon approche sera limitée à la résolution de

¹ *La recherche*, n° 348, rubrique "SAVOIRS"

l'hétérogénéité énonciative de premier degré (entre les informateurs), parce que j'ai volontairement limité mon objectif à la controverse entre les thèses scientifiques sur le gène : il faudrait reprendre l'analyse au point de départ (et non ajouter après-coup l'analyse du chapeau introductif et des schémas et dessins) si l'on voulait embrasser méthodiquement l'ensemble de l'article.

Mon hypothèse tient en peu de mots : les positions de chacun ne peuvent être comprises que:

1* si on a identifié la nature et l'enjeu de cette divergence entre les dix-huit définitions, à savoir la tension difficile à résoudre entre deux grands types de définitions du gène ;

2* si on a pu reconstituer les différents cas de figure de la relation entre ces deux définitions (notamment : ascendante / descendante ; causale / statistique ; disjointe / conjointe ; avec transition / sans transition ; isotope / allotope)

3* quand on pu situer chaque position par rapport à ce dispositif global, qui constitue l'*univers de discours* des dix-huit textes analysés.

Les deux définitions du gène

Ces deux définitions sont :

- 1- le gène comme séquence moléculaire matérielle
- 2- le gène comme unité d'information

Du point de vue de la mise en scène textuelle, le traitement de ces deux définitions est franchement contrasté.

LE GÈNE COMME SÉQUENCE MOLÉCULAIRE MATÉRIELLE

La séquence moléculaire matérielle est définie en extension. Il s'agit de définir ses constituants, ses limites, ses formes alternatives ; les constituants sont définis dans une séquence de production :

(1) la *régulation* : des "régions" d'ADN sont des régulateurs pour les opérations qui vont être accomplies sur la région d'ADN dite "ADN promoteur" ;

* la question se pose donc de savoir si ces régions régulatrices font ou ne font pas partie de la séquence ADN pertinente ;

(2) la *transcription* : l'ADN promoteur est "suivi" de manière plus ou moins complète et régulière par l'ARN dit "ARN transcripteur" (une enzyme polymérase), pour le transcrire, et produire un autre ARN, dit "ARN transcrit" ou "ARN messenger" ;

* les décisions à prendre se multiplient :

- l'ARN transcripteur est-il ou pas un constituant de la séquence matérielle à retenir ?
- comme certaines régions de l'ADN promoteur ne sont pas transcrites (les segments dits "introns") font-ils partie ou pas de la séquence pertinente ?
- comme le début et la fin de la séquence transcrite sont indiqués par deux nucléotides, des "codons", respectivement le "codon start" et le "codon stop", ces deux codons font-ils partie ou pas de la séquence pertinente ?

(3) la *synthèse* : l'ARN transcrit et un acide aminé sont assemblés grâce à une molécule intermédiaire, un ARN dit "ARN de transfert", et on peut alors dire que l'ARN transcrit a "synthétisé" une protéine ;

* l'identification des constituants de la séquence biologique originelle pourrait s'arrêter à la phase précédente, puisqu'il n'y a maintenant plus aucune trace de l'ADN originel, mais le questionnement continue, dans la mesure où, dès qu'on a renoncé à limiter la séquence moléculaire pertinente à l'ADN promoteur, on ne peut plus ignorer les bases moléculaires de la suite des opérations : ainsi se pose la question de savoir si l'ARN transfert fait partie ou pas de cette séquence.

(4) la *maturation* : la protéine synthétisée subit de nombreuses modifications sous l'effet de son environnement chimique ;

* si on pousse la logique de ce raisonnement jusqu'au bout, tout ce qui modifie la nature du produit des opérations antérieures doit entrer dans la définition en extension du gène, et par conséquent, les facteurs de modification de la protéine en font partie.

On remarque tout de suite dans cette présentation trois types de phénomènes :

a) Si l'on se fixe strictement sur la séquence biologique originelle du gène, la définition s'arrête au point (2), et plus précisément à l'ADN promoteur, délimité par deux bornes indiquées

par les deux codons.

Dès qu'on prend en compte d'autres éléments, notamment les régions régulatrices, et les divers ARN (transcripteur, messenger et transfert), on change de perspective, puisqu'il s'agit alors d'un processus de production et non d'un état matériel originel. Le simple fait de s'interroger sur la pertinence, au sein même de la séquence d'ADN promoteur, des segments qui sont ignorés par la transcription (les introns), constitue un raisonnement rétroactif, qui s'appuie sur le processus de production, et non sur la seule délimitation de la séquence matérielle originelle.

b) L'identification de ce "saut qualitatif" n'est possible que si on a préalablement identifié l'existence des phases du processus de production, et ordonné les définitions en fonction de ces quatre phases. La seule lecture des dix-huit définitions, prises séparément, ne permet pas en effet de déceler un tel saut : la discussion y apparaît simplement comme une recherche des constituants biologiques du gène, et semble porter seulement sur leur nombre, sur l'extension de la définition. Mais la notion d'extension, en l'occurrence, n'a pas grand sens, puisque ces différents éléments ne constituent pas une "classe" homogène (des segments d'ADN + plusieurs types d'ARN) : le gène n'est donc pas, à ce stade de la réflexion, et dans la perspective de l'identification de la séquence biologique pertinente, à proprement parler une "classe".

c) En d'autres termes, la discussion sur le nombre de composants matériels du gène repose sur *un micro-récit sous-jacent*, celui de la production des chaînes moléculaires biologiques : le gène est une "chaîne" de séquences biologiques qui s'engendrent mutuellement, qui s'emboîtent et se combinent selon un processus explicite. Mais cette "chaîne" n'étant définissable qu'à partir du processus, *le gène est le processus*, et pas seulement l'addition des différentes molécules ou segments de molécules qui y participent.

On hésiterait donc ici entre une définition "segmentale" et une définition "supra-segmentale", entre une définition par les "figures" et une définition par les "processus". Mais c'est la définition processuelle qui l'emporte, car ceux qui, comme Fischer et Jordan, veulent limiter la définition à la séquence ADN originelle, adoptent une argumentation restrictive par rapport au processus : Fischer, par exemple, rappelle le processus (le codage des transcrits, et la synthèse des protéines) mais pour préciser que le gène n'est que la séquence sur laquelle opère la première étape du processus.

LE GÈNE COMME UNITÉ D'INFORMATION

La définition ne vise plus une séquence biologique, mais une fonction biologique, et elle va se présenter comme une formulation d'actes et de mises en corrélation entre des phénomènes.

Le gène “code” pour la synthèse de protéines.

Le gène est une “instruction” pour un programme génétique.

Le gène “détermine” l’apparition d’un caractère phénotypique.

La diffusion d’un gène est “statistiquement corrélée” à la diffusion d’un phénotype.

Le gène “conditionne” deux propriétés essentielles et indissociables du vivant : sa reproductibilité et sa variabilité.

Cette série de définitions appelle elle aussi plusieurs commentaires.

1- On retrouve ici un processus, mais saisi globalement, comme une opération sur des informations, des instructions, et pas le processus de production des structures matérielles. La production des structures matérielles visait à déterminer les composants en extension du gène, pour en faire le “produit” d’une suite d’opérations, alors que le processus “informatif” s’empare du gène comme un tout, pour en faire un opérateur, dont les produits sont des observables dans le monde vivant.

Le changement de statut du gène est double :

- d’un point de vue actantiel, il passe du statut de produit à celui de producteur
- d’un point de vue modal, il passe d’un statut “pragmatique” et “expliqué”, un composé de figures matérielles descriptives, à un statut “cognitif” et explicitement sémiotique, ou du moins “explicatif” : *marqueur, instruction, programmation, déterminant.*

2- Le niveau de compréhension et de comparabilité des définitions dépend de la procédure adoptée : saisies chacune séparément, ces définitions apparaissent comme des variations, plus ou moins précises et assumées, sur un même thème ; mais, saisies comparativement et globalement, elles peuvent être elles-mêmes ordonnées selon deux paramètres :

- le statut modal de l’explication ;
- la portée de l’explication.

a) Le statut modal de l’explication varie en effet depuis quelque chose qui s’apparenterait à une “causalité”, jusqu’à une simple corrélation statistique : en somme, en suivant la gamme des degrés de probabilité et de certitude, les définitions parcourent l’axe des modalités épistémiques, depuis le “certain” (le gène “code”, fournit des “instructions” pour un programme) jusqu’au simplement “probable” (le gène “détermine”, “conditionne”, est “corrélé statistiquement”).

b) La portée de la relation et de l’opération varie de son côté en suivant le processus même de l’engendrement des chaînes matérielles du vivant : depuis les protéines (le gène “code” la synthèse des protéines), jusqu’au vivant en général (le gène conditionne la

reproductibilité et la variabilité du vivant), en passant par les phénotypes (le gène détermine l'apparition de caractères phénotypiques).

Mais seule cette mise en perspective globale permet de comprendre que le processus en question, dont le gène est l'opérateur "cognitif" n'est pas de même nature que le processus de production matérielle exploité dans l'autre définition du gène : en effet, quand on adopte la définition du gène comme "unité d'information", on ne peut plus "suivre à la trace", étape par étape, la production de la chaîne matérielle. Les deux paramètres de variation de la définition, le paramètre modal et le paramètre de "portée", introduisent des sauts qualitatifs irréversibles : la corrélation statistique, notamment, interdit de rabattre la relation sur une chaîne de produits matériels ; de même, la reproductibilité et la variabilité du vivant ne peut pas se réduire à une combinaison de phénotypes, pas plus qu'un phénotype ne peut se réduire à une combinaison de protéines.

Les relations entre les deux définitions

LE RÔLE DE LA QUANTIFICATION

A ce point de l'analyse, il devient particulièrement ardu de rendre compte de l'articulation entre les deux définitions, puisqu'on en est à constater une série de "sauts qualitatifs irréversibles", et un "rabattement impossible" de l'une sur l'autre.

L'analyse séparée de chacune des définitions ne pourra que constater la divergence de traitement de cette articulation :

- certains n'adoptent qu'une seule définition, en général la définition matérielle, et récusent ou ignorent l'autre ;
- certains constatent la divergence entre les deux définitions, et même la radicalisent ("L'ADN n'est pas porteur du gène");
- d'autres enfin engagent une démarche de réduction de la divergence, et tentent l'articulation: nous reviendrons sur ces tentatives.

Mais il est au moins un point sur lequel tous sont d'accord (ou, au moins, ne manifestent pas de désaccord) : la difficulté vient de la "*quantification*". Une des définitions repose même entièrement sur la quantification, celle de Chambon : ce n'est pas l'unité de transcription qui porte le gène, mais le *nombre* de protéines synthétisées par le génome.

Or la quantification est, dans cet ensemble de définitions, une propriété attachée au processus de production matériel, et son effet ne peut être apprécié que si on a reconstitué les

étapes et les opérations de ce processus ; on a donc affaire à une syntaxe d'opérations, et la propriété critique de cette syntaxe est son *aspectualité*, et plus précisément, la *quantification* qui va s'appliquer et se combiner à cette aspectualité : il faut avoir préalablement fait l'analyse aspectuelle du processus pour repérer les zones critiques de la quantification. Or l'effet de la quantification, très précisément, est la modification du statut modal et de la portée du processus même : au fur et à mesure où on avance dans la chaîne de production des figures matérielles, la quantification des opérations, par un effet cumulatif, interdit la rétroaction causale, et oblige à faire le "saut cognitif".

Voilà les "zones critiques" du processus :

(1) Il y a toujours sur l'ADN plusieurs régions de fixation possibles pour l'ARN polymérase : le choix du promoteur est donc multiple.

(2) Pour le même promoteur ADN, il y a toujours plusieurs bornes ou sites de début et de fin : le choix de la séquence codante précise est donc lui aussi multiple.

- plusieurs sites d'initiation possibles (plusieurs points de fixation du codon "start")

- plusieurs sites de terminaison possibles (plusieurs points de fixation du codon "stop").

(3) Lors de la transcription, le phénomène de l'épissage alternatif va lui aussi multiplier les possibilités de production de protéines : au cours du processus de duplication, l'ARN polymérase qui suit la séquence ADN de référence peut dupliquer les deux brins de l'hélice à la fois, un seul d'entre eux, ou, le plus souvent, sauter de l'un à l'autre, selon un rythme et une fréquence particulière : il y a donc là aussi un très grand nombre de solutions possibles.

On note donc quatre facteurs de pluralisation de la seule opération de transcription, qui résultent de la quantification des seuils aspectuels du processus de production matérielle : (1) le nombre de promoteurs, (2) la position du site initial, (3) la position du site terminal, (4) la stratégie de l'épissage alternatif. A chaque choix pour chacune de ces variables, correspond un ARN "transcrit" différent, et, par conséquent, une protéine différente.

Mais il faut ajouter à cela que, pour chacune des protéines, et à partir d'un même transcrit (ARN messenger), on obtient encore des devenir différents, selon l'ARN de transfert qui sert d'interface, et selon les conditions de maturation de la protéine : cela porte donc à six le nombre de phases (aspects) et donc à six le nombre de bifurcation par quantification : choix du promoteur (*fixation*), choix du site d'initiation (*start*), choix du site de terminaison (*stop*), choix des brins de l'hélice (*épissage alternatif*), choix de l'ARN de transfert (*collage*), devenir de la protéine (*maturation*).

On comprend alors pourquoi la *quantification des aspects* du procès fait difficulté, et pourquoi, au moment de l'interprétation et de la tentative de définition, c'est elle, précisément, qui change le statut du processus de production des protéines et des phénotypes. La raison en est simple: le nombre de possibilités, pour chaque choix, n'est pas déterminé, et le nombre de choix à faire aboutit globalement à une indétermination de la lecture ascendante du processus.

Comment expliquer cette "conversion cognitive" ?

Chacun des facteurs de pluralisation est en même temps *un facteur de "tri"* : s'il y a plusieurs solutions, et si une seule est réalisée à la fois, alors il faut supposer *une instance de contrôle* (le rôle des régions régulatrices n'est pas clairement défini dans le corpus). Et, aussi longtemps que les fonctions de "tri" ne peuvent pas être rapportées à des productions chimiques spécifiques, elles apparaissent déjà comme des fonctions cognitives de "contrôle" ; certes, au moins dans un cas, on identifie la présence de segments moléculaires qui jouent le rôle de "codons" (opérateurs pragmatiques du tri), mais on ne précise pas pourquoi, où et comment sont produits ces "codons" ; à la place de ces explications, les définitions font appel aux "régions de régulation", c'est-à-dire à *des opérateurs "cognitifs" du tri*.

RELATION ASCENDANTE ET RELATION DESCENDANTE

La quantification introduit donc une dissymétrie entre les deux directions du parcours :

- dans le sens ascendant, se produit un saut modal et un changement du statut modal du processus de production du vivant, et, par conséquent, du gène lui-même ;
- dans le sens descendant, ce saut modal ne se produit pas, puisqu'on part des observables (les phénotypes) et que, par analyse, on retrouvera les séquences moléculaires qui les déterminent.

1- *Dans le sens ascendant,*

le pragmatique est "converti" en cognitif.

L' "information" est alors une catégorie qui apparaît quand la relation de causalité n'est plus linéaire et univoque ; on parle d'autant plus d' "information" et de "code" qu'on maîtrise moins la relation entre le référent et le produit : par exemple, quand on cherche à passer du génotype au phénotype, le saut est plus important qu'entre l'ADN promoteur et la protéine, et l'explication par l' "information" et les modalités cognitives (probabilités, incertitudes) devient dominante. Dans le sens ascendant, la quantification des opérations pragmatiques provoque la *conversion sémiotique* : la quantification du processus pragmatique implique des opérations de tri, qui supposent un contrôle cognitif.

Ce phénomène est proprement textuel (et même narratif), et n'est pas spécifique au discours scientifique sur le gène. Il est en effet comparable à la reconstruction de la signification d'un récit: elle ne peut être que rétrospective, car dans le sens prospectif, l'information contenue dans chaque phase ou périphérie peut produire un grand nombre de bifurcations.

En sémiotique narrative, on distingue ainsi deux "directions méthodologiques" : la direction "ascendante", qui ne débouche que sur le constat de la multitude des "possibles narratifs", et sur l'identification des "points de bifurcation" ; la direction "descendante", pour laquelle, rétrospectivement, chaque point de bifurcation apparaît après coup comme contrôlé par une modalité (un *devoir faire*, un *vouloir faire*, un *savoir faire* ou un *pouvoir faire*). En d'autres termes, ce qui apparaît comme "quantification des possibles" dans le sens ascendant sera traité comme "modalisation du processus" dans le sens descendant. En effet, s'il y avait une relation univoque entre chaque phase narrative, et un enchaînement sans bifurcation, il n'y n'aurait besoin d'aucun savoir-faire, vouloir-faire ou pouvoir-faire, la conversion d'une phase à l'autre serait automatique et directe.

Ce point pourrait paraître inutilement technique et purement analogique, mais il n'en est rien, car il permet de montrer très précisément que le problème qui se pose dans cette collection de définitions :

- est un problème de nature sémiotique, qui concerne la dimension narrative de la mise en scène discursive des phénomènes évoqués,
- et qu'il est trans-textuel ; en effet, il n'est pas localisable dans telle ou telle définition, et seule la confrontation et la mise en ordre de toutes les définitions permet de le faire apparaître ; il s'agit d'un phénomène discursif, commun à l'ensemble des définitions, et non pas au sens où l'ensemble de ces définitions formeraient un "corpus", une "classe" de textes analysables dans leurs récurrences et leurs différences, mais au sens où chacun participe d'une syntaxe englobante, y occupe sa place, en décrit une des étapes, etc.

2- *Dans le sens descendant,*

toutes les définitions s'accordent sur ce point : il n'y a pas de relation cognitive envisageable ; on ne sait pas passer des phénotypes au génotype, et du génotype aux séquences moléculaires, par des opérations cognitives ; il y a des corrélations, mais établies statistiquement sur des bases de données, et pas sur des observations suivies de prévision. On peut observer, identifier, isoler, etc. le gène correspondant à un phénotype ou une protéine, on ne peut pas le prévoir, le calculer, se le représenter à partir de la protéine.

Dans le sens descendant, on opère donc, en revanche, une "*conversion*" pragmatique, qui consiste à repérer expérimentalement et par observation directe les unités matérielles (les

structures biologiques) qui sont supposées à l'origine de la production des unités observables : à partir d'un phénotype, on peut retrouver la séquence d'ARN, puis à partir de celle-ci, on peut identifier la séquence ADN de référence.

Dans cette conversion pragmatique de l' "unité d'information", la quantification s'inverse, car s'il y a plusieurs phénotypes possibles issus d'un même gène, à partir d'un ou plusieurs phénotypes donnés, on peut remonter au gène unique qui lui ou leur correspond.

CONCLUSIONS

Une démarche qui aurait consisté à étudier chaque définition séparément et à tenter ensuite de les confronter n'aurait finalement rien pu montrer d'autre que ce qu'on constate intuitivement immédiatement à la lecture : les généticiens ne sont pas d'accord entre eux.

La démarche adoptée, qui se voulait d'emblée intégrative, partait du postulat que ces différentes définitions partageaient un même univers de discours, même si elles se présentaient comme un corpus hétérogène. Si on admet, dans la perspective de la théorie polyphonique et de l'intertextualité, qu'on n'énonce jamais seul et pour la première fois, mais que, chaque fois qu'on tient un discours, il est en relation avec une multitude d'autres discours, effectifs, antérieurs ou potentiels, il faut bien que ce principe ne s'applique pas seulement à l'analyse des textes particuliers et isolés, mais que, au premier chef, il s'applique aux ensembles textuels qui partagent une même thématique, un même genre, où quelque forme sémiotique que ce soit.

Quel que ce soit le cas de figure, l'énonciation individuelle n'est qu'une spécification, une actualisation et une restriction à partir de la praxis énonciative collective. Dans le cas qui nous occupe, ce principe est en outre conforté par le fait que les personnes interrogées font partie d'une même communauté scientifique, connaissent chacun, vraisemblablement, les positions des uns et des autres, ou, au moins, les positions-types qui existent dans cette communauté : il est toujours plus confortable de proposer une lecture intertextuelle entre René Char et les Présocratiques si on peut s'assurer que René Char a lu les Présocratiques. On a donc affaire à une véritable structure conversationnelle à distance. Mais l'existence d'une communauté professionnelle ou institutionnelle n'est qu'un cas particulier de ces "praxis énonciatives collectives" et de ces dialogues involontaires qui se nouent dans les corpus hétérogènes.

Et c'est justement cette démarche qui a permis de mettre en lumière, en deux étapes successives :

1) La problématique et les propriétés des deux définitions concurrentes : à ce niveau d'analyse, chaque définition pourrait (c'est ce qui sera fait dans une version plus longue de cette

étude) être caractérisée comme une position d'équilibre (ou de déséquilibre) dans la structure tensile qui associe et confronte à la fois le gène comme structure matérielle et le gène comme structure d'information.

2) La problématique de la conversion et de la relation entre les deux grands types de définition : ce ne sont pas vraiment deux objets sémiotiques différents, mais un seul objet, qui reçoit deux types de conversions. Dans le sens ascendant, une *conversion cognitive*, qui fait de l'objet "moléculaire" une "unité d'information", et, dans le sens descendant, une *conversion pragmatique*, qui fait du phénotype et des résultats de la sélection naturelle, l'expression d'une structure moléculaire de référence. Ces deux "conversions", ascendantes et descendantes, correspondent respectivement :

- pour la *conversion ascendante*, à la reconnaissance des *contenus cognitifs* associés aux *expressions biologiques* ;

- pour la *conversion descendante*, à l'affectation d'une *expression matérielle* aux *contenus d'information*.

On assiste donc à un processus de sémiotisation orienté :

- dans le sens ascendant, ce processus de sémiotisation "invente" en quelque sorte des "contenus" d'information, des codages et des opérations de tri et de sélection ;

- dans le sens descendant, le processus de sémiotisation "ancree" les phénomènes cognitifs et informatifs dans la matière, les inscrit sur des supports biologiques, et leur trouve en quelque sorte des "expressions" matérielles.

On a donc affaire à deux sémiotiques différentes et associées par une fonction : d'un côté une *sémiotique modale et cognitive*, qui associe des "indices", des "expressions" à des contenus d'information, et à la formation de systèmes de valeurs ; de l'autre une sémiotique des "modus opérandi", ce qu'on appellerait aujourd'hui une "*sémiotique de l'empreinte*", c'est-à-dire une sémiotique qui focaliserait sur les modalités de l'inscription sur un support matériel de l'expression, aux formes de l'interaction avec ce support et à la syntaxe propre de ce dernier.

L'une de ces sémiotiques donne la prééminence à *la génération du plan des contenus* (sélection, valeurs, adaptation, devenir du vivant, etc.). L'autre donne la prééminence à *la génération du plan des expressions* (support, inscription, syntaxe, modus opérandi, etc.). La réunion de ces deux macro-sémiotiques forme une *sémiotique-objet complète et biplane*, à partir de laquelle, justement parce qu'elle est biplane, on peut tenir un raisonnement, dégager de l'intelligibilité, assurer une compréhension, etc.

ANNEXES : résumé des 18 définitions

1) H. Atlan

Propose quatre définitions pouvant être regroupées en trois entrées complémentaires :

- a) -séquence d'ADN pouvant coder plusieurs protéines
- b) b1 -lieu d'action de la sélection naturelle (évolution), mais la sélection ne porte que sur les phénotypes, et pas sur l'ADN
- c) c1 -instruction de programme génétique
- c2 - séquence d'ADN associée statistiquement à un trait de caractère

Les deux premières définitions appartiennent à deux domaines différents, et actualisent donc deux “isotopies” distinctes : la définition chimique et moléculaire, et la définition “évolutionniste”. La troisième met en relation les deux premières : entre les deux rôles de l'ADN (structure biologique moléculaire & ressort de la sélection), il y a un “code”, un “programme”, ou une corrélation statistique.

La définition installe progressivement une contre-définition, puisque chaque proposition est corrigée et annulée par la précédente. Atlan finit même par dire que tout cela n'est qu'une “simplification idéologique et journalistique”, puisque la séquence d'ADN est un marqueur (c'est-à-dire un signe) et qu'on ne peut pas établir de relation de causalité entre l'ADN et les phénotypes de la sélection naturelle, et ce d'autant plus que ce n'est pas l'ADN qui est actif, mais les ARN transcrits et les protéines synthétisées).

2) P. Chambon

P. Chambon retient les mêmes éléments de définition (molécule, codage), mais ne mentionne pas la question des phénotypes et de la sélection. Il ajoute en outre le premier rôle de l'ADN (unité de transcription).

Mais surtout, il introduit le thème de la quantification :

- plusieurs ADN “promoteurs” possibles pour une même unité de transcription
- plusieurs sites d'initiation et de terminaison pour une même unité de transcription (produisant plusieurs “transcrits)
- plusieurs protéines possibles à partir de chaque transcrit

Ce qui lui permet de passer de l' “unité” au tout : ce n'est pas l'ADN-unité de transcription qui

porte le gène, mais l'ensemble et le nombre des protéines codées par le génome (100000 chez la mouche, 1 million chez l'homme, pour une même base de 30000 gènes)

3) D. Cohen

D. Cohen tient le même raisonnement, concernant l'absence de causalité directe, et le rôle de la quantification ; il rajoute même deux motifs supplémentaires :

- (1) épissage séparé ou alternatif des deux brins de la séquence d'ADN, pouvant coder séparément ou alternativement pour des protéines différentes, et
- (2) les régions régulatrices de la transcription peuvent être plus ou moins éloignées de la séquence codante.

Mais au lieu de déplacer la question vers la statistique ou vers le nombre global, il insiste lui aussi sur le rôle de l'ARN (les transcrits, et les régulateurs de la transcription), qui seuls ont un rapport avec l'individualité de chacun, à travers les protéines qu'ils synthétisent.

Lui aussi suscite le doute aussitôt qu'il a avancé une solution (polymorphisme de l'ARN)

4) A. Fischer

A. Fischer ramène la définition au noyau biologique et moléculaire : un gène est une séquence d'ADN, qui code des ARN, lesquels synthétisent des protéines.

5) Ph. Froguet

Ph. Froguet déplace clairement la définition du côté des phénotypes et de la sélection naturelle : le gène est pour lui une unité déterminant un caractère phénotypique, et en conséquence, il propose de déplacer aussi l'intérêt du chercheur vers les phénomènes "aval" (le métabolome, la maturation et la structure des protéines dans leur environnement).

Il a donc renoncé à puiser les éléments de sa définition dans la structure bio-chimique, ainsi que dans les questions de codage ou de programmation ("déterminant" laisse une grande marge d'appréciation).

6) J.-Ch. Galloux

En tant que juriste, J.-Ch. Galloux doit prendre en compte les deux dimensions du gène, mais sans se préoccuper de leur mise en relation : pour lui, en effet, ce sont deux entités juridiquement indépendantes, et qui relèvent de deux domaines différents du droit :

(1) d'un côté la structure bio-chimique dans sa matérialité, qui ancre le gène dans la catégorie des choses que l'on peut s'approprier, et qui relèvent donc du droit de la propriété privée ;

(2) de l'autre l'unité d'information, le déterminant du phénotype, qui range le gène dans la catégorie des "informations privées" et c'est à ce titre qu'il relève des droits immatériels de la personne.

7) P. Gaudray

P. Gaudray reprend en détail les éléments de la description biologique déjà connue, en ajoutant quelques autres facteurs de complexification et de quantification : on se demande alors si, dans une définition en extension, il faudrait intégrer à ce qu'on appelle un "gène" les parties d'ADN éliminées et non transcrites (les introns), les ARN traducteurs (ribosomes), les ARN de transfert (en interface entre l'ARN transcrit et l'acide aminé dans la protéine).

Il insiste sur la dissymétrie du raisonnement, qui explique à la fois la croyance (en un lien fort, en une corrélation) et l'absence de démonstration causale :

- on connaît les gènes à partir des phénotypes (sens descendant), mais
- on ne connaît pas les phénotypes à partir des gènes (sens ascendant)

8) J. Générmont

J. Générmont introduit une notion nouvelle, qui permet d'aborder différemment la relation entre la structure moléculaire et son rôle génétique : il s'agit des allèles de la séquence d'ADN ; au lieu d'être définis à partir de leur structure biologique, ils sont définis par deux propriétés fonctionnelles : l'isoaction (c'est une classe de séquences ADN produisant le même phénotype) et l'homologie (c'est une classe qui présente un nombre suffisamment de traits d'organisation communs).

Les séquences biologiques matérielles sont à cet égard de simples "motifs contextuels", déterminés par leur position sur un chromosome.

9) P.-H. Gouyon

P.-H. Gouyon reprend la définition par les allèles en la précisant : à l'intérieur d'un ensemble de gènes homologues, c'est-à-dire susceptibles de s'apparier, et formant un "locus", une partie d'entre eux donnent le même phénotype : ils sont donc "isoactifs". C'est cet ensemble de gènes homologues et isoactifs qui définissent des allèles (les motifs différents de ces gènes homologues isoactifs)

Le raisonnement cherche visiblement à ménager des phases de transition entre la structure

biologique matérielle et la structure d'information associée à la sélection naturelle, en combinant deux réductions, une ascendante (l'homologie définit une classe de séquences appariables, un *locus*) et une autre descendante (un phénotype donné pouvant correspondre à un grand nombre de séquences ADN, il présuppose donc une classe d' *isoaction*, la classe de toutes les séquences pouvant être associées à ce phénotype).

La réduction ascendante et la réduction descendante se rencontrent alors pour imposer une double réduction définitionnelle à l'allèle : une définition morphologique (réduction ascendante) et une définition fonctionnelle (réduction descendante).

A la fois classe morphologique et classe isoactive ("prédicative"), l'allèle est un actant: il n'y a plus besoin de se demander à son propos si, comment et pourquoi il engendre un phénotype, puisque telle sa définition ; il n'y a plus besoin de se demander si le responsable est bien telle séquence ADN, puisque la séquence ADN n'est qu'une réalisation locale et contrainte de l'allèle.

10) Ph. Jarne

Ph. Jarne reprend la plupart des motifs précédents, en insistant sur le caractère statistique de la relation entre génotype et phénotype, mais aussi sur le caractère aléatoire de la relation entre allèle et motif moléculaire.

11) B. Jordan

Définition purement moléculaire, focalisée sur la séquence ADN, et tout particulièrement sur les marqueurs de début et de fin de la partie transcrite, ainsi que sur les autres segments (de régulation, et non transcrits).

12) A. Khan

A. Khan choisit de radicaliser la duplicité du gène, en fournissant la composition chimique de la séquence ADN (polymère formé de nucléotides (molécules formées de base+sucre+acide phosphorique), et en généralisant la définition fonctionnelle à hauteur du "vivant en général" : le gène conditionne la vie, ses particularités et son évolution (membrane + ensemble fini + métabolisme + reproductibilité et variabilité codée par les gènes).

C'est donc, eu égard à la double définition avec laquelle la plupart cherchent à composer ou à biaiser, l'écart maximal : non seulement il n'y a plus de relation entre les deux, mais en outre, chacune des définitions est formulée de manière à être entièrement allotope par rapport à l'autre.

13) A. Laganey

Se contente de constater le nombre illimité de définitions, et reprend la notion d' "unité d'information transmise".

14) D. Louvard

Dans la relation entre les deux définitions, introduit une variante (contrôle d'un phénotype), mais pour le reste, utilise les mêmes éléments (unité d'information transmissible mais altérable)

15) Fr. Rehenmann

Tout comme Jordan, donne une définition purement moléculaire et focalisée sur les limites de la séquence pertinente : pour cela, il isole les "codons" (suite de trois nucléotides) pour les fonctions "Start" (initiation de la transcription) et "Stop" (terminaison de la transcription) : un gène est la séquence définie par ces deux bornes.

Revient lui aussi sur la multiplicité des définitions, et sur le caractère statistique de la relation explicative.

16) P. Sonigo

Reprend l'idée de la double définition (informative, permettant le "calcul" d'hérédité, & moléculaire et matérielle), mais pour insister sur leur irréductible incompatibilité (le lien entre le gène et le caractère observable, bilatéral dans la première définition, est indéterminé et purement statistique dans la deuxième.

Sa conclusion est donc radicale : l'ADN ne porte pas de gènes !

17) L. Thaler

- idem : tout ADN n'est pas formé de gènes

18) J. Warcoin

Définition purement biologique, et focalisée sur la détermination des parties pertinentes: du point de vue de la propriété industrielle, il faut surtout délimiter ce qui appartient au gène (par ex. les introns qui sont détruits au moment de la transcription, ou les régions de régulation)